PNAS
www.pnas.org

Supplementary Information Appendix for

RUNX1 AND RUNX3 DRIVE PROGENITOR TO T-LINEAGE TRANSCRIPTOME CONVERSION IN T-CELL COMMITMENT VIA DYNAMIC GENOMIC SITE SWITCHING

Boyoung Shin[1]*, Hiroyuki Hosokawa[1,2]*, Maile Romero-Wolf[1]†, Wen Zhou[1], Kaori Masuhara[2], Ditsa Levanon[3], Yoram Groner[3], Ellen V. Rothenberg[1]‡

‡Corresponding author: Ellen V. Rothenberg

Email: evroth@its.caltech.edu

This PDF file includes:

Supplementary text

List of Supplemental Datasets

Figures S1 to S12

SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S5

Supplementary Information Text

Materials and Methods

Mice

C57BL/6J, B6.Cg-Tg(BCL2)25Wehi/J(Bcl2-tg) and B6.Gt(ROSA)26Sortm1.1(CAG-cas9*,-EGFP)Fezh/J (Cas9) mice were purchased from the Jackson Laboratory and bred at the California Institute of Technology. B6. *Bcl11b*$^{yfp/yfp}$ reporter mice and B6. *Bcl11b*$^{mCh/mCh}$ reporter mice were described previously (1, 2). Both male and female mice were used at ages from 8-12-week-old for bone marrow and at 4-6 weeks old for thymus samples. All animals were bred and maintained under specific pathogen-free conditions at the California Institute of Technology according to Institutional Animal Care and Use Committee (IACUC) regulations.

Cell Lines

The Scid.adh.2c2 pro-T cell-like cell line (3) was cultured in RPMI1640 with 10% fetal bovine serum (FBS, Sigma-Aldrich), sodium pyruvate (Gibco), non-essential amino acids (Gibco), Pen-Strep-Glutamine (Gibco) and 50 mM β-mercaptoethanol (Sigma-Aldrich).

An ILC2 cell line, ILC2/b6 (4) was cultured in OP9 medium (α-MEM, 20% FBS, 50 mM β-mercaptoethanol, Pen-Step-Glutamine) supplemented with 10 ng/ml of IL-2, IL-7 and IL-33 (Pepro Tech Inc.). Transduction of cell lines and western blotting were carried out as described (5).

In vitro cell culture

Bone marrow was obtained from the femurs and tibiae of 8-12 week-old C57BL/6J mice or progeny of Cas9 x Bcl2-tg mice or progeny of Cas9 x Bcl2-tg x Bcl11b$^{mCh/mCh}$ mice. Progenitor cells from the bone marrow cell suspension were enriched by depleting mature lineage^{+} cells expressing CD3ε (clone 145-2C11), CD19 (clone 1D3), B220 (clone RA3-6B2), NK1.1 (clone PK136), CD11b (clone M1/70). CD11c (clone N418), Gr1 (clone RB6-8C5), and Ter119 (clone TER-119) using MACS LS magnetic columns (Miltenyi Biotec). Enriched progenitor cells were co-cultured with OP9-Dll1 or OP9-Mig cells (6), which were obtained from Dr. Zúñiga-Pflücker. During in vitro culture, cells were supplemented with IL-7 (Peprotech) and Flt3l (Peprotech) at various concentrations: 10 ng/ml each from day0 – day7, 5 ng/ml each from day7- day10, and 1 ng/ml each from day10 – day18 in OP9 medium (α-MEM, 20% FBS, 2mM glutamine, 100 IU/ml penicillin, 100 mg/ml streptomycin, and 50 μM β-ME). All in vitro cultures were done under 37 °C, 7% CO_2 environment. Detailed methods were described previously (2, 7, 8).

Specific guide RNA design and transduction

The 19-mer sgRNAs against Runx1, Runx3, and a luciferase control were designed and delivered as described previously (7, 9). The CHOPCHOP web tool (https://chopchop.rc.fas.harvard.edu/) was used to design the sgRNAs. The designed sgRNAs were inserted into the empty E42 dTet vector, which expresses human U6 promoter and a reporter molecule (either mTurquoise2 or human NGFR) as described previously (7). In order to assure deletion of the target gene, three different sgRNA expressing vectors were generated each for Runx1 and for Runx3, and each trio was pooled for packaging in Phoenix-Eco cells using Fugene 6 Transfection Reagent (Roche) as described previously (10, 11).

For transductions, non-tissue culture plates were coated with 50 μg/ml RetroNectin (Takara bio) at 4°C overnight. After removal of excess RetroNectin, viral supernatant was added to the plate and centrifuged at 2000×g, 32°C for 2 hours. Then pro-T cells from OP9-Dll1 culture on day2 (for Phase1) or on day10 (for Phase2) were detached from the OP9-Dll1 stromal cells and incubated

with retroviral vectors bound plate for 4 hours at 37°C. During 4 hours of infection, IL-7, Flt3l, and SCF (10 ng/ml each for Phase1, 1ng/ml each for Phase2) were supplemented. After infection, cells were removed from the viral particles and cultured with OP9-Dll1 or OP9-Mig cells for 3-7 additional days.

Sequences for sgRunx1 and sgRunx3 are below.

Control (Luciferase) ggcatttcgcagcctaccg

sgRunx1 #1 gctcgtgctggcatctacg

sgRunx1 #2 agccccggcaagatgagcg

sgRunx1 #3 agcggcgaccgcagcatgg

sgRunx3 #1 tacgggaatacgcatgccg

sgRunx3 #2 taagcgcgcaggcaaccgc

sgRunx3 #3 gaagcgttgcgcagctcgg

Flow cytometry analysis and cell sorting

Cell surface staining was performed following Fc blocking by incubating single cell suspensions in 2.4G2 hybridoma cell supernatant. Then cells were stained with a biotin-conjugated lineage cocktail (TCRβ (ebioscience, clone H57-597), TCRγδ (eBioscience, clone GL-3), CD19, NK1.1, CD11b, CD11c, and Gr1). Secondary surface staining was performed with fluorescently conjugated streptavidin, CD45, cKit (clone 2B8), CD44 (clone IM7), CD25 (clone PC61.5), and hNGFR (clone ME20.4). A viability dye (Life Technologies, Aqua) or 7AAD (eBioscience) was applied to exclude dead cells.

For intracellular staining for Runx1 and Runx3, cells were fixed with 3.7% formaldehyde for 15 min at room temperature after surface staining. Then cells were permeablized with the Foxp3

Permeabilization/Fixation kit (eBioscience) and stained with fluorescently conjugated with Runx1(clone: RXDMC) and Runx3(clone: 527327).

Samples were acquired using MACSQuant 10 (Miltenyi Biotec) and CytoFlex (Beckman Coulter) and data was analyzed with FlowJo version 10.6.1 (Tree Star).

For RNA-seq on sgRNA introduced pro-T cells, bone marrow progenitor cells from Cas9;Bcl2 animals were subjected to in vitro culture as described. On day3 post infection Phase1 and Phase2 cells were sorted for Lin^- $CD45^+$ $mTurquoise2^+$ (marker for sgControl or sgRunx3) $hNGFR^+$ (marker for sgControl or sgRunx1) population. In order to obtain reference RNA-seq for supervised PCA analysis, bone marrow progenitor cells were obtained from B6.$Bcl11b^{yfp/yfp}$ reporter animals and cultured with OP9-Dll1 cells for 4-6 days. ETP, Bcl11b/YFP^- DN2a, and Bcl11b/YFP^+ DN2b cells were sorted based on cKit, CD44, and CD25 expression as described previously (12). For ChIP-seq, cells were sorted for Lin^- $CD45^+$ $cKit^{high}$ $CD25^-$ population on day 5 of culture (Phase1) or Lin^- $CD45^+$ $cKit^{low}$ $CD25^+$ cells on day 14 of culture (Phase2) using BD FACSAria and BD FACSAria Fusion at the California Institute of Technology Flow Cytometry Facility.

RNA purification and RNA sequencing

For RNA sequencing, total RNA was isolated from 50,000 ~ 250,000 cells using the RNeasy Micro Kit (QIAGEN) according to the manufacturer's protocols. Sequencing libraries were constructed at the California Institute of Technology Genomics Facility using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB) and sequenced on Illumina HiSeq2500 (1x50 base pair, single-read).

Chromatin Immunoprecipitation (ChIP) and deep sequencing

ChIP-seq was performed as described previously (5, 7, 8). Briefly, $1x10^7$of Phase1 (on day5) or Phase2 (on day14) cells were sorted from *in vitro* culture and crosslinked with 1mg/ml DSG

(Thermo Scientific) followed by 1% formaldehyde. The reaction was quenched by 0.125M glycine. Nuclei were isolated by incubating crosslinked cells in Nuclei Isolation buffer (50 mM Tris-pH 8.0, 60 mM KCl, 0.5% NP40) and lysed in Lysis buffer (0.5% SDS, 10 mM EDTA, 0.5 mM EGTA, 50 mM Tris-HCl (pH 8)). The lysates were sonicated on a Bioruptor (Diagenode) for 18 cycles (one cycle: 30sec max power sonication followed by 30 sec rest). Rabbit anti-Runx1 antibody (Abcam, 5 µg per $1x10^7$cells) or Rabbit anti-Runx3 antibody (13) was bound to Dynabeads anti-Rabbit (Invitrogen) and incubated with sonicated chromatin in 1X RIPA buffer at 4°C overnight. After washes, precipitated chromatin fragments were eluted in ChIP elution buffer (20 mM Tris-HCl, pH 7.5, 5 mM EDTA 50 mM NaCl, 1% SDS, and 50 µg proteinase K) by incubating at 65°C for 6 hours. Eluted DNA was cleaned up using Zymo ChIP DNA Clean & Concentrator according to manufacturers' protocols. ChIP-seq libraries were prepared using NEBNext ChIP-Seq Library Preparation Kit (NEB) and sequenced on Illumina HiSeq2500 (1x50 base pair, single-read) at the California Institute of Technology Genomics Facility.

Immunoblotting

A Cas9-GFP expressing DN3-like cell line, Scid.adh.2c2, and a Cas9-GFP expressing ILC2 cell line, ILC2/b6, were infected with sgRunx1-hNGFR or sgRunx3-CFP. Four days after sgRNA transduction, sgRunx1 introduced GFP^+hNGFR^+ Scid.adh.2c2 cells, and sgRunx3 introduced GFP^+CFP^+ ILC2/b6 cells were sorted and subjected to immunoblotting. Cytoplasmic and nuclear extracts, used to the detection of Tubulina, and Runx1 and Runx3, respectively, were prepared using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Pierce). Lysates were run on 10 % polyacrylamide gel, followed by immunoblotting. The antibodies used for the immunoblot analysis were anti-Tubulina (Sigma, T6199), anti-Runx1 (Abcam) and anti-Runx3 (Levanon et al., 2014).

Resource details

Reagent	Source	Identifier
Anti-human/mouse CD44 PE (clone IM7)	eBioscience	Cat#12-0441-83
Anti-mouse CD117 (cKit) APC (clone 2B8)	eBioscience	Cat#17-1171-82
Anti-mouse CD25 APCe780 (clone PC61.5)	eBioscience	Cat#47-0251-82
Anti-mouse CD45 PECy7 (clone 30-F11)	eBioscience	Cat#25-0451-82
anti-human NGFR PE (clone ME20.4)	eBioscience	Cat#12-9400-42
anti-human NGFR PECy7 (clone ME20.4)	BioLegend	Cat#345109
Anti-mouse TCRβ Biotin (clone H57-597)	eBioscience	Cat#13-5961-85
Anti-mouse TCRγδ Biotin (clone GL-3)	eBioscience	Cat#13-5711-85
Anti-mouse CD3ε Biotin (clone 145-2C11)	eBioscience	Cat#13-0031-82
Anti-mouse CD8α Biotin (clone 53-6.7)	eBioscience	Cat#13-0081-86
Anti-mouse CD19 Biotin (clone 1D3)	eBioscience	Cat#13-0193-85
Anti-mouse B220 Biotin (RA3-6B2)	eBioscience	Cat#13-0452-85
Anti-mouse Gr1 (clone RB6-8C5)	eBioscience	Cat#13-5931-86
Anti-mouse NK1.1 Biotin (clone PK136)	eBioscience	Cat#13-5941-85
Anti-mouse CD11b Biotin (clone M1/70)	eBioscience	Cat#13-0112-86
Anti-mouse CD11c Biotin (clone N418)	eBioscience	Cat#13-0114-85
Anti-human/mouse Runx1 PE (clone RXDMC)	eBioscience	Cat#12-9816-80
Anti-human/mouse Runx3 Alexa Fluor 647 (clone 527327)	R&D systems	Cat#C3765R-100UG
Rabbit polyclonal anti-Runx1	Abcam	Cat#ab23980
Rabbit polyclonal anti-Runx3	Levanon et al., 2014	N/A
α-MEM	GIBCO	Cat#12561-056
Fetal Bovine Serum	SigmaAldrich	Cat#F7305
β-mercaptoethanol	SigmaAldrich	Cat#M6250
Penicillin-Streptomycin-Glutamine (100X)	Gibco	Cat#10378-016
HBSS	GIBCO	Cat#14175-095
HEPES	GIBCO	Cat#15630-080
RetroNectin	Takara	Cat#T100B

FuGENE 6 Transfection Reagent	Promega	Cat#E2691
37% formaldehyde	ThermoFisher Scientific	Cat#28906
DSG (disuccinimidyl glutarate)	ThermoFisher Scientific	Cat#20593
Complete, EDTA-free protease inhibitor cocktail	Roche	Cat#11873580001
Dynabeads M-280 Sheep anti-Rabbit IgG	Invitrogen	Cat#11204D
Proteinase K Solution	ThermoFisher Scientific	Cat#AM2548
7AAD	eBioscience	Cat#00-6993-50
MACS Streptavidin Microbeads	Miltenyi Biotec	Cat#130-048-101
Human IL-7	PeproTech Inc	Cat#200-07
Human FLT-3-Ligand	PeproTech Inc	Cat#300-19
Human Stem Cell Factor	PeproTech Inc	Cat#250-03
Murine M-CSF	PeproTech Inc	Cat#315-02
Mouse GM-CSF	Miltenyi Biotec	Cat#130-095-739
Murine IL-6	PeproTech Inc	Cat#216-16
MACS LS columns	Miltenyi Biotec	Cat#130-042-401
LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit	Life Technologies	Cat#L34966
Foxp3/Transcription Factor Staining Buffer Set	eBioscience	Cat#00-5523-00
Illumina Nextera DNA preparation Kit	Illumina	Cat#FC-121-1030
Nextera Index Kit (96 indexes, 384 samples)	Illumina	Cat#FC-121-1012
RNeasy Micro Kit	QIAGEN	Cat#74004
ChIP DNA Clean and Concentrator	Zymo Research	Cat#D5205
NEBNext ChIP-Seq Library Preparation Kit	NEB	Cat#NEB #E6240
NEBNext Ultra RNA Library Prep Ki	NEB	Cat#NEB #E7530
High Sensitivity DNA Kit	Agilent Technologies	Cat#5067- 4626
Qubit dsDNA HS Kit	ThermoFisher Scientific	Cat#Q32854
SPRIselect reagent kit	Beckman Coulter	Cat#B23318
Agencourt AMPure XP beads	Beckman Coulter	Cat#A63880

RNA-seq analysis

After the base calling, each sample had approximately 30 million read depth per sample. The adapter trimmed reads were aligned to the mouse reference genome GRCm38/mm10 using STAR (v2.4.0) and gene expression was calculated with RSEM(v1.2.25) by following the ENCODE RNA-seq pipelines. Significant changes in transcript expression were determined using EdgeR (v.3.6.8). Differentially expressed genes (DEGs) were defined from the resulting normalized datasets by FPKM $\geq$ 1, adjusted p-value < 0.05, and $\log_2$fold-change > 0.5. Non-DEGs were defined by FPKM $\geq$ 1, adjusted p-value $\geq$ 0.05. For creating heatmaps, FPKM normalized reads of DEGs were $\log_2$ transformed and hierarchically clustered with R hclust function (Manhattan distance, complete linkage). Gene expression data was visualized as $\log_2$(FPKM+0.1) using R pheatmap. Volcano plot and scatter plots were generated in Prism8 software (v 8.4.3)(GraphPad).

For gene set enrichment analysis (GSEA), RNA sequencing data was pre-ranked according to an adjusted p value and $\log_2$ fold-change. The gene sets from the hallmark gene sets (H), curated gene sets (C2), regulatory target gene set (C3), gene ontology (GO) gene sets (C5), and immunologic signature gene sets (C7) of the Molecular Signatures Database (MsigDB) were used for computing enrichment. The normalized enrichment score (NES), nominal p value, and false detection rate (FDR) q-value were assessed using GSEA software (GSEA (v.4.0.3) from Broad Institute by running in pre-ranked list mode with 1,000 permutations. The significant pathways obtained by FDR q value < 0.05 were visualized using Python Matplotlib.

For supervised Principal Component Analysis (PCA), the PC loadings for T developmental trajectory genes was obtained from a previous single cell RNA-seq study (12). Briefly, curated regulatory genes were determined based on functional importance during thymic T development (www.immgen.org)(14), reviewed in (15, 16)). After the PC analysis on scRNA-seq, the PC loadings of principal components 1 and 2 were projected to bulk RNA-seq data. Reference bulk RNA-seq results for *in vivo* thymocytes were reported previously (7, 12). Reference bulk RNA-seq for *in vitro* data was performed by culturing bone marrow progenitor cells with OP9-Dll1 cells in OP9 media

supplemented with IL-7 and Flt3l for 4-6 days. Following populations were sorted and subjected to bulk RNA-seq: ETP (cKit^{high} CD25^{-} Bcl11b/YFP^{-}), DN2a Bcl11b- (cKit^{high} CD25^{+} Bcl11b/YFP^{-}), and DN2a Bcl11b^{+} (cKit^{high} CD25^{+} Bcl11b/YFP^{+}).

ChIP-seq analysis

Sequenced reads (approximately 30 million reads per sample) were mapped to the mouse reference genome GRCm38/mm10 using Bowtie (v1.1.1) and reproducible peak calling was performed using a HOMER (17) adaptation of the Irreproducibility Discovery Rate (IDR) tool according to ENCODE guideline. For downstream analysis, peaks with a normalized peak score ≥ 15 were considered.

Peak overlaps between different ChIP-seq samples were computed using a HOMER package (mergePeak.pl) and the resulted groups were visualized as heat maps or area proportional Venn diagrams or scatter plots. Peak centered heat maps were created with a HOMER package (annotatePeaks.pl -ghist) in a 2000 bp region and hierarchical clustering was performed using Cluster3. TreeView (v.1.1.6r4) was used to visualize heat maps. Area proportional Venn diagrams were generated using Python Matplotlib-venn tools (v.0.11.5). Scatter plot was generated by counting tag densities from indicated tag directories. The resulting tag counts per 10 million reads (base 2 logarithmic converted) were visualized using Python holoviews (v.1.13.3) with datashader (v.0.9.0) operation.

Motif enrichment analysis was performed in the HOMER package (findMotifGenome.pl) using a 200bp window and *De novo* results were reported. Peaks were annotated to genomic regions using HOMER package (annotatePeaks.pl).

For UCSC Genome Browser visualization, BigWig files were generated from the aligned SAM or BED-file formats using Samtools and Bedtools, and then normalized to 1 million reads.

Runx1 and PU.1 ChIP-seq data in Phase1, and Runx1 and Bcl11b ChIP-seq data in Phase2 used in this study are previously published (7, 18, 19)

RNA-seq and ChIP-seq association analysis

Genes associated with ChIP-seq peaks were annotated using GREAT (v.4.0.4) with proximal: 5kb upstream, 1kb downstream, plus distal: up to 1000kb mode (20). Non-promoter peaks with a normalized peak score $\geq$ 30 were annotated. Promoter-associated peaks were not included for this analysis because they elevated the background of binding unrelated to function, an issue especially important to control for under the ChIP-seq crosslinking conditions used here. To analyze association between peak-annotated genes and Runx DEGs, the count table for each ChIP-seq group and RNA-seq clusters was generated and independence was determined by Fisher's exact test. Fisher's exact test was performed with Monte Carlo simulation to compute *P*-values. For the groups with *P*-value < 0.001, residual analysis was performed. Fisher's exact test and standardized residuals calculation was performed using R (version 3.6.1).

ATAC-seq data analysis

Publicly available ATAC-seq data (GSE100738) was downloaded as raw sequence read files and mapped onto GRCm38/mm10. After filtering out mitochondrial reads, peak calling was performed using Genrich (ver. 0.6) and Motif enrichment was computed using the same settings as described above. The frequency of motif occurrence in regions of 2000 bp surrounding the ATAC sites were analyzed by using a HOMER package (annotatePeaks.pl -size <#> -hist <#> -m <motif file>). The resulting histograms were visualized using Python bokeh plotting (v.2.1.1).

The Gene Expression Omnibus (GEO) accession numbers for RNA-seq, ChIP-seq, and ATAC-seq analysis are presented below.

Deposited Data	Reference	GEO number
Phase1 and Phase2 Runx1 ChIP-seq	Hosokawa et al., 2018b (18)	GSE103953
Phase1 PU.1 ChIP-seq	Ungerback et al., 2018 (8); Zhang et al., 2012 (19)	GSE93755; GSE31235
Phase2 Bcl11b ChIP-seq	Hosokawa et al., 2018a (7)	GSE110305
ETP, Bcl11b- DN2a, and Bcl11b+ DN2a RNA-seq	Zhou et al., 2019 (12)	GSE130812
DN2b and DN3 RNA-seq	Hosokawa et al., 2018a (7)	GSE115744
ATAC-seq	Yoshida et al., 2019 (21)	GSE100738
RNA-seq and ChIP-seq	This paper	GSE154304

Other statistical tests

To compare the average of three or more groups, One-way ANOVA with Tukey's multiple comparisons test was used. To assess linear correlation between two different ChIP-seq peaks or gene expression fold-changes, Pearson correlation coefficient (Pearson's *r*) was calculated. The best-fit line with 95% confidence interval was obtained by linear regression. Nonparametric tests comparing two distributions were performed by two-sample Kolmogorov-Smirnov test. To compare two population proportion, *Z*-test was performed. Fisher's exact test and standardized residual analysis was conducted to evaluate association between categorical variables from ChIP-seq and RNA-seq groups. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$ for One-way ANOVA. *P*-values for Kolmogorov-Smirnov test are denoted on the figures. * |*z*-score| > 1.9599; ** |*z*-score| > 2.5758; *** |*z*-score| > 3.2905 for standardized residual analysis. One-way ANOVA was performed using Prism software (version 8.4.3, GraphPad). Two-sample Kolmogorov-Smirnov test was computed using scipy.stats.ks_2samp (version 1.4.1) and *Z*-test was computed using statsmodels.stats.proportion

(version 0.11.1) from Python (version 3.7.7). Pearson's *r* calculation, linear regression analysis, Fisher's exact test, and standardized residual analysis were performed using R (version 3.6.1).

List of Supplemental Datasets

1. Supplemental Dataset S1: Enriched motifs at pro-T cell ATAC sites and Runx ChIP-seq peaks

2. Supplemental Dataset S2: seqFISH transcript counts for *Runx* family gene expression in individual thymic DN cells

3. Supplemental Dataset S3: *Runx* deletion-sensitive differentially expressed genes (DEGs)

4. Supplemental Dataset S4: PCA coordinate loadings for reference and *Runx* knockout samples in supervised PCA analysis

5. Supplemental Dataset S5: Table of *Runx*-sensitive DEG – Runx ChIP peak association values

SI References

SUPPLEMENTARY REFERENCES

1. Ng KK, *et al.* (2018) A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. *Elife* 7:e37851.

2. Kueh HY, *et al.* (2016) Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. *Nat Immunol* 17:956-65.

3. Dionne CJ, *et al.* (2005) Subversion of T lineage commitment by PU.1 in a clonal cell line system. *Dev Biol* 280:448-466.

4. Zhang K, *et al.* (2017) Cutting Edge: Notch Signaling Promotes the Plasticity of Group-2 Innate Lymphoid Cells. *J Immunol* 198:1798-1803.

5. Hosokawa H, *et al.* (2020) Cell type-specific actions of Bcl11b in early T-lineage and group 2 innate lymphoid cells. *J Exp Med* 217:e20190972.

6. Schmitt TM & Zúñiga-Pflücker JC (2002) Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. *Immunity* 17:749-756.

7. Hosokawa H, *et al.* (2018) Bcl11b sets pro-T cell fate by site-specific cofactor recruitment and by repressing *Id2* and *Zbtb16*. *Nat Immunol* 19:1427-1440.

8. Ungerbäck J, *et al.* (2018) Pioneering, chromatin remodeling, and epigenetic constraint in early T-cell gene regulation by SPI1 (PU.1). *Genome Res* 28:1508-1519.

9. Romero-Wolf M, *et al.* (2020) Notch2 complements Notch1 to mediate inductive signaling that initiates early T cell development. *J Cell Biol* 219:e202005093.

10. Del Real MM & Rothenberg EV (2013) Architecture of a lymphomyeloid developmental switch controlled by PU.1, Notch and Gata3. *Development* 140:1207-1219.

11. Champhekar A, *et al.* (2015) Regulation of early T-lineage gene expression and developmental progression by the progenitor cell transcription factor PU.1. *Genes Dev* 29:832-848.

12. Zhou W, *et al.* (2019) Single-cell analysis reveals regulatory gene expression dynamics leading to lineage commitment in early T cell development. *Cell Syst* 9:321-337 e9.

13. Levanon D*, et al.* (2014) Transcription factor Runx3 regulates interleukin-15-dependent natural killer cell activation. *Mol Cell Biol* 34:1158-69.

14. Mingueneau M*, et al.* (2013) The transcriptional landscape of αβ T cell differentiation. *Nat Immunol* 14:619-632.

15. Yui MA & Rothenberg EV (2014) Developmental gene networks: a triathlon on the course to T cell identity. *Nat Rev Immunol* 14:529-545.

16. Longabaugh WJR*, et al.* (2017) Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. *Proc Natl Acad Sci U S A* 114:5800-5807.

17. Heinz S*, et al.* (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576-589.

18. Hosokawa H*, et al.* (2018) Transcription Factor PU.1 Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding. *Immunity* 48:1119-1134 e7.

19. Zhang JA, Mortazavi A, Williams BA, Wold BJ, & Rothenberg EV (2012) Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* 149:467-482.

20. McLean CY*, et al.* (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28:495-501.

21. Yoshida H*, et al.* (2019) The cis-Regulatory Atlas of the Mouse Immune System. *Cell* 176:897-912 e20.

A.

ATAC peaks in ETP

Motif	% of targets % of Background	p-vlaue TF
	30.73% 7.88%	1e-8958 Spi1
	13.15% 1.80%	1e-6286 CTCF(Zf)
	20.67% 7.86%	1e-3296 Runx(Runt)
	10.09% 6.20%	1e-451 Foxo1

ATAC peaks in DN2a

Motif	% of targets % of Background	p-vlaue TF
	33.22% 12.08%	1e-5975 ELF3(ETS)
	8.21% 0.71%	1e-5067 BORIS(Zf)
	20.11% 7.24%	1e-3375 Runx(Runt)
	2.63% 0.82%	1e-497 NF-E2(bZIP)

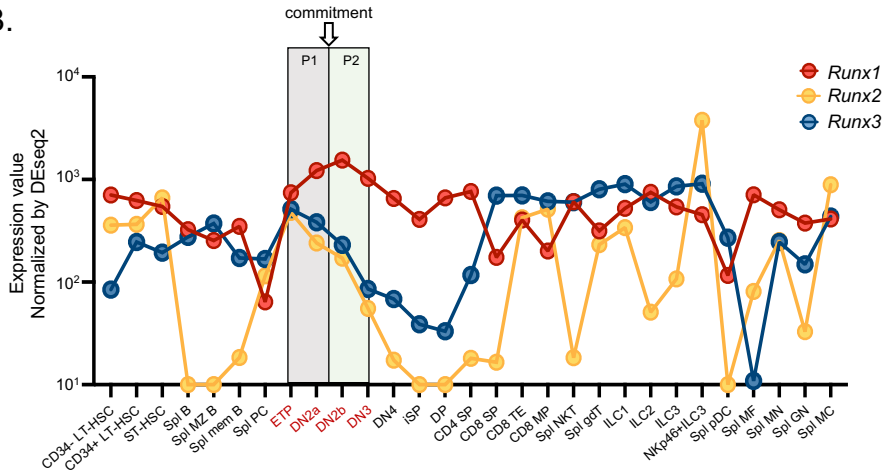
ATAC peaks in DN2b

Motif	% of targets % of Background	p-vlaue TF
	9.56% 0.61%	1e-4791 BORIS(Zf)
	32.00% 12.77%	1e-3347 ETS1(ETS)
	14.06% 4.33%	1e-1969 Runx(Runt)
	26.44% 17.32%	1e-696 Ptf1a(bHLH)

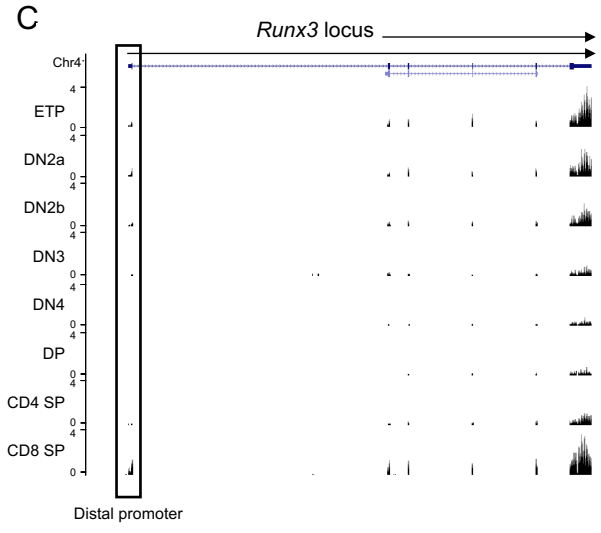
ATAC peaks in DN3

Motif	% of targets % of Background	p-vlaue TF
	9.87% 0.50%	1e-3255 BORIS(Zf)
	21.87% 8.30%	1e-1373 ERG(ETS)
	11.76% 4.20%	1e-769 Runx(Runt)
	26.21% 16.74%	1e-451 E2A(bHLH)

B.



C



D.

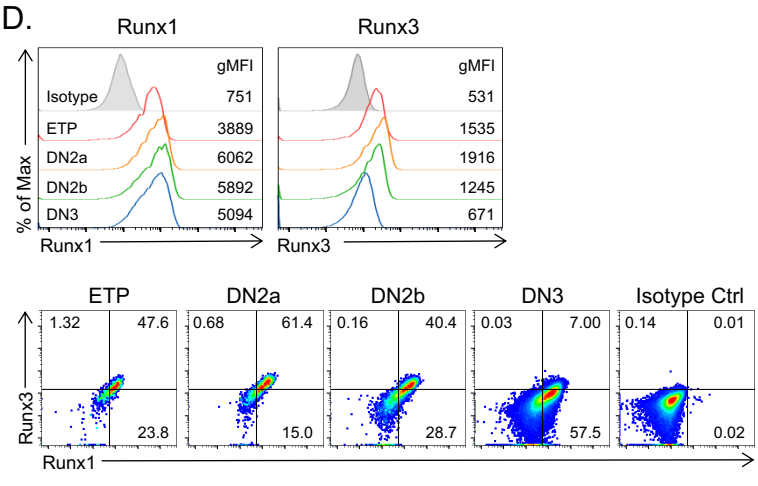
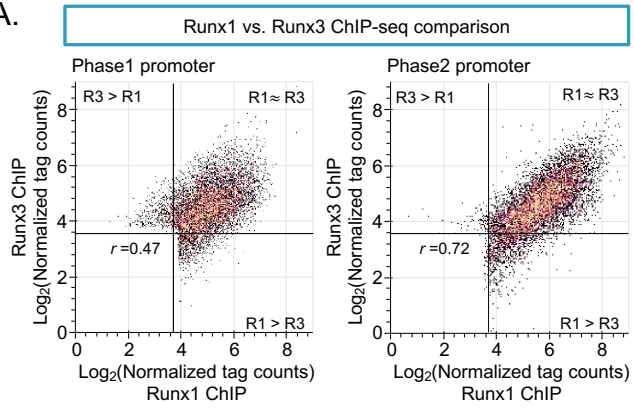


Figure S1.

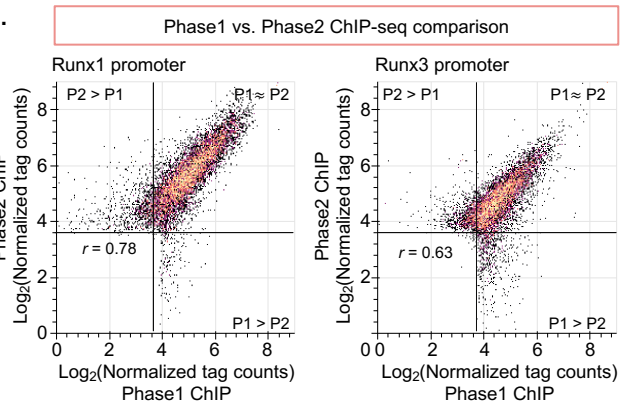
Fig. S1. Runx factors are highly expressed in pro-T cells and their motif sequences are enriched in open chromatin regions of pro-T cells.

(A) Top four enriched sequence motifs of open chromatin regions from ETP, DN2a, DN2b, and DN3 ATAC peaks are shown. (B) Graph shows expression profiles of *Runx1*, *Runx2*, and *Runx3* in hematopoietic lineage cells. P1; Phase1, P2; Phase2. (C) UCSC genome browser tracks show *Runx3* gene locus with arrows indicating transcripts from distal and proximal promoters and RNA-seq tracks from indicated thymocyte subpopulations. Box encloses exon 1 of *Runx3* transcripts initiated from the distal promoter, showing clear expression in ETP-DN2b stages as well as well-known expression in CD8 SP. (D) Runx1 and Runx3 protein levels in WT thymic DN populations were measured by flow cytometry. Representative plots were gated on live Lin^- cells and subpopulations were determined by cKit and CD25 expression: ETP ($cKit^{high}$ $CD25^-$), DN2a ($cKit^{high}$ $CD25^+$), DN2b ($cKit^{int}$ $CD25^+$), and DN3($cKit^{low}$ $CD25^+$) (n = 9). Geometric MFI (gMFI) and frequency are reported. (A-C) Data from immgen.org (21).

A.



B.



C.

Runx1 non- promoter
Phase1 and Phase2 shared

Motif	% of targets % of Background	p-vlaue TF
	43.99% 10.13%	1e-1429 Runx(Runt)
	44.54% 10.74%	1e-1387 ERG(ETS)
	6.68% 0.82%	1e-315 CTCF(Zf)
	8.43% 3.14%	1e-121 JunB(bZIP)

Runx1 non-promoter
Phase1 specific

Motif	% of targets % of Background	p-vlaue TF
	38.90% 2.32%	1e-1547 Spi1
	42.09% 18.03%	1e-298 Runx(Runt)
	9.41% 1.28%	1e-211 BORIS(Zf)
	1.12% 0.00%	1e-112 FOXH1

Runx1 non-promoter
Phase2 specific

Motif	% of targets % of Background	p-vlaue TF
	49.45% 12.32%	1e-2964 Runx2(Runt)
	31.81% 6.72%	1e-1850 ERG(ETS)
	34.64% 14.73%	1e-807 E2A(bHLH)
	25.16% 13.39%	1e-327 LEF1(HMG)

D.

Runx3 non-promoter
Phase1 and Phase2 shared

Motif	% of targets % of Background	p-vlaue TF
	64.51% 15.25%	1e-3167 Runx2(Runt)
	46.43% 12.60%	1e-1750 ERG(ETS)
	14.53% 4.80%	1e-352 IRF4(IRF)
	3.11% 0.52%	1e-153 BORIS(Zf)

Runx3 non-promoter
Phase1 specific

Motif	% of targets % of Background	p-vlaue TF
	47.63% 8.70%	1e-2368 Spi1
	49.23% 9.60%	1e-2337 Runx2(Runt)
	1.92% 0.02%	1e-312 XCPE1
	1.57% 0.02%	1e-261 AR-halfsite(NR)

Runx3 non-promoter
Phase2 specific

Motif	% of targets % of Background	p-vlaue TF
	49.44% 10.10%	1e-1708 Runx2(Runt)
	27.85% 5.56%	1e-880 ERG(ETS)
	36.60% 17.09%	1e-376 Ptf1a(bHLH)
	21.90% 9.11%	1e-255 TCF7L2

E.

Phase1 non-promoter
Runx1 and Runx3 shared

Motif	% of targets % of Background	p-value TF
	46.60% 10.03%	1e-1988 Spi1
	55.23% 15.04%	1e-1963 Runx(Runt)
	7.72% 2.59%	1e-158 AP-1(bZIP)
	2.81% 0.45%	1e-130 CTCF

Phase1 non-promoter
Runx1 > Runx3

Motif	% of targets % of Background	p-value TF
	26.92% 1.68%	1e-594 BORIS(Zf)
	18.32% 1.98%	1e-290 SpiB(ETS)
	18.24% 8.71%	1e-51 GC box
	4.82% 1.10%	1e-40 Nrf2(bZIP)

Phase1 non-promoter
Runx3 > Runx1

Motif	% of targets % of Background	p-value TF
	47.48% 5.93%	1e-3444 Runx(Runt)
	41.40% 11.25%	1e-1497 Spi1
	1.34% 0.01%	1e-261 ZBTB7C
	1.14% 0.01%	1e-198 Tcfap2c

F.

Phase2 non-promoter
Runx1 and Runx3 shared

Motif	% of targets % of Background	p-value TF
	44.62% 7.75%	1e-3480 Runx1(Runt)
	48.00% 13.17%	1e-2473 ETS1(ETS)
	28.86% 14.80%	1e-451 E2A(bHLH)
	20.02% 10.86%	1e-251 TCF7L2

Phase2 non-promoter
Runx1 > Runx3

Motif	% of targets % of Background	p-value TF
	27.25% 6.19%	1e-740 ERG(ETS)
	42.30% 17.65%	1e-557 Runx2(Runt)
	35.94% 17.65%	1e-321 HEB(bHLH)
	7.29% 1.05%	1e-275 BORIS(Zf)

Phase2 non-promoter
Runx3 > Runx1

Motif	% of targets % of Background	p-value TF
	68.60% 18.04%	1e-931 Runx(Runt)
	23.33% 6.39%	1e-226 ERG(ETS)
	4.48% 0.51%	1e-90 BORIS(Zf)
	12.87% 5.07%	1e-69 Tcf7(HMG)

Figure S2.

Fig. S2. Genomic interaction patterns of Runx1 and Runx3 in promoter regions are similar, whereas non-promoter Runx peaks show distinct motif enrichments at different developmental stages.

(A, B) Scatter plots display $\log_2$ normalized tag counts per 10 million tags from promoter regions with Pearson correlation *r*. Runx1 vs. Runx3 comparison (A), and Phase1 vs. Phase2 comparison (B). (C, D) Top four enriched sequence motifs found at ChIP-seq defined binding sites of Runx1 (C) and Runx3 (D) in different T-cell developmental phases are shown. (E, F) Top four enriched sequence motifs in different groups of Runx peaks at Phase1 (E) or Phase2 (F) are shown.

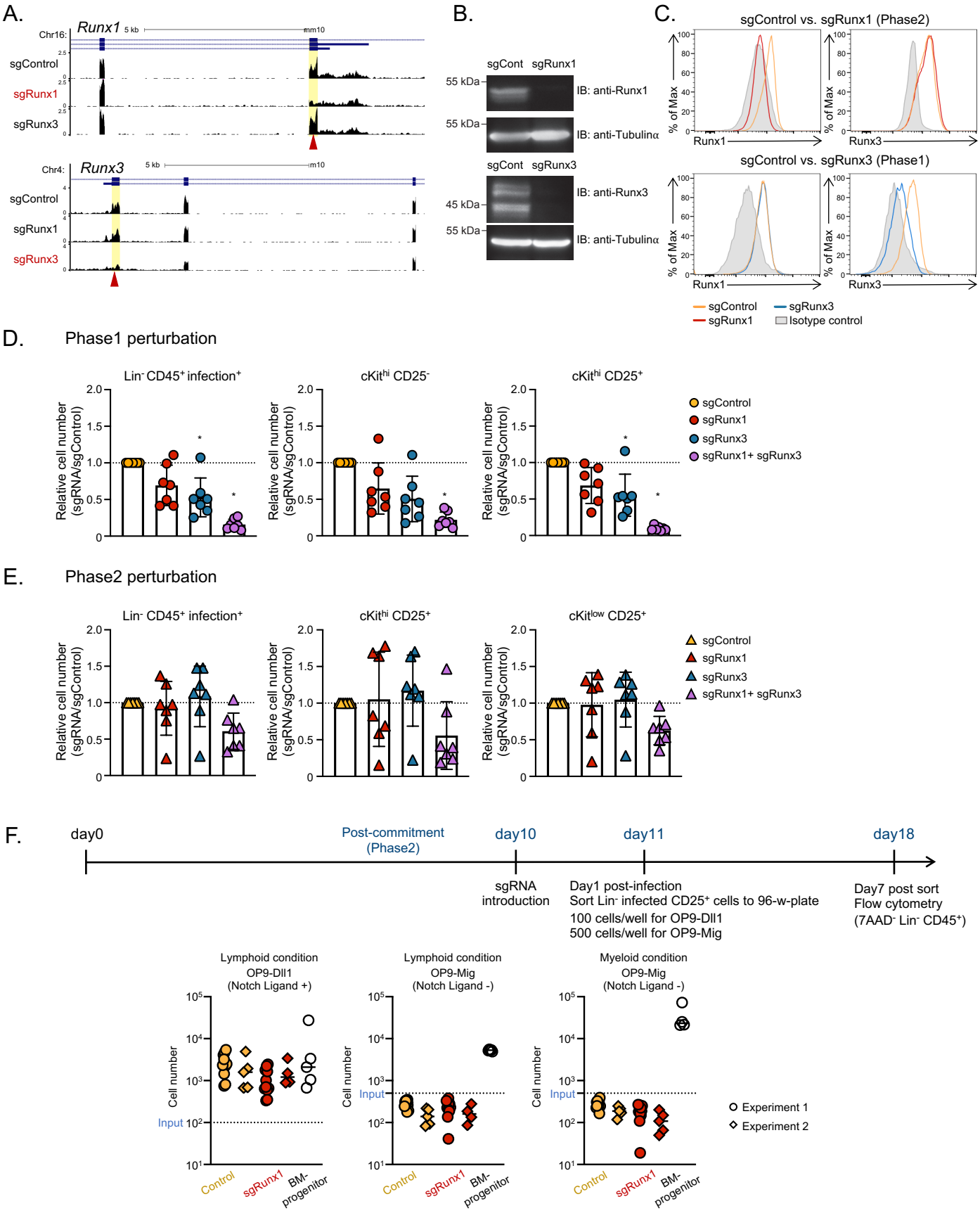

A.
Runx1
Chr16:
sgControl
sgRunx1
sgRunx3
Runx3
Chr4:
B.
sgCont sgRunx1
55 kDa
IB: anti-Runx1
IB: anti-Tubulinα
sgCont sgRunx3
45 kDa
IB: anti-Runx3
C.
sgControl vs. sgRunx1 (Phase2)
sgControl vs. sgRunx3 (Phase1)
% of Max
Runx1
Runx3
sgControl
sgRunx1
sgRunx3
Isotype control
D.
Phase1 perturbation
Lin- CD45+ infection+
cKithi CD25-
cKithi CD25+
Relative cell number (sgRNA/sgControl)
sgControl
sgRunx1
sgRunx3
sgRunx1+ sgRunx3
E.
Phase2 perturbation
Lin- CD45+ infection+
cKithi CD25+
cKitlow CD25+
F.
day0
Post-commitment (Phase2)
day10
day11
day18
sgRNA introduction
Day1 post-infection
Sort Lin- infected CD25+ cells to 96-w-plate
100 cells/well for OP9-Dll1
500 cells/well for OP9-Mig
Day7 post sort
Flow cytometry
(7AAD- Lin- CD45+)
Lymphoid condition
OP9-Dll1
(Notch Ligand +)
Lymphoid condition
OP9-Mig
(Notch Ligand -)
Myeloid condition
OP9-Mig
(Notch Ligand -)
Cell number
Input
Control
sgRunx1
BM-progenitor
Experiment 1
Experiment 2

Figure S3.

Fig. S3. CRISPR-Cas9 system deletes Runx factors effectively and shows contribution of Runx factors regulating cell recovery in Phase1 and Phase2.

Bone marrow progenitor from Cas9;Bcl2 mice or Cas9;Bcl2 mice with Bcl11b-mCherry reporter were cultured with OP9-Dll1 cells for 2 days (Phase1) or 10 days (Phase2). sgRNAs against Runx1 and/or Runx3 were introduced retrovirally and infected cells were co-cultured with OP9-Dll1 cells for 3 days before analysis. (A) Effects of Cas9-mediated deletion on transcribed sequences from *Runx1* and *Runx3*, in zoomed-in views of *Runx1* and *Runx3* loci showing regions targeted by sgRNA (arrowheads). These regions are in exon 2 of each gene, which is the first exon used in common by transcripts initiating at both proximal and distal promoters. Cas9 and sgRNA cleavages induce nested sets of small deletions focused around the sites complementary to the sgRNA, giving specific depletion of RNA-seq reads from the targeted regions. Effects of targeted deletion of Runx1 (top) and Runx3 (bottom) are displayed in the same RNA samples used to define the Runx-sensitive DEGs (yellow highlighted regions with red arrow). (B) Western blots showing effective loss of Runx1 and Runx3 protein after Cas9-mediate deletion, using Scid.adh.2c2 and ILC2/b6 cell lines (5). Four days after sgRNA transduction, lysates from retrovirus infected GFP^+hNGFR^+ Scid.adh.2c2 cells and GFP^+CFP^+ ILC2/b6 cells were subjected to immunoblotting for Runx1 and Runx3, respectively. Three independent experiments were performed with similar results. (C) Expression levels of Runx1 and Runx3 proteins were measured by flow cytometry in Phase1 pro-T cells (for sgRunx3 conditions) and Phase2 (for sgRunx1 conditions) pro-T cells 3-4 days after sgRNA introduction. Plots representative of four independent experiments are gated on Lin^- $CD45^+$ infection marker $(CFP\&NGFR)^+$ populations. Note that samples were analyzed with different flow cytometer settings for Phase1 than for Phase2 samples and levels of staining can only be compared for samples in the same Phases. (D,E) Role of Runx1 and Runx3 in sustaining pro-T cell populations. Relative cell numbers of indicated populations compared to sgControl transduced cells for Phase1 (D) and Phase2 (E) are shown (n =7). Graphs show the average ± S.D. (F) Deletion

of Runx1 after Phase1 to Phase2 transition does not reverse T-lineage commitment. Bone marrow progenitor cells from Cas9;Bcl2 animals were driven through T-cell lineage commitment with OP9-Dll1 cells for 10 days and then infected with sgControl or sgRunx1. One day post infection, $7AAD^-$ Lin^- $CD45^+$ infection marker $(NGFR)^+$ $cKit^-$ $CD44^-$ $CD25^+$ cells were sorted to a 96-well-plate and cultured with either OP9-Dll1 cells, to continue T-cell development, or OP9-Mig cells to force conversion to an alternative fate (100 input cells per well for OP9-Dll1 culture and 500 input cells per well for OP9-Mig culture). Cells were supplemented with lymphoid condition cytokines or myeloid condition cytokines. As a positive control for lineage plasticity, BM progenitor cells without any T-lineage promoting preculture were incubated in parallel and shown to grow well in all three conditions. After 7 days, viability of Lin^- $CD45^+$ cells was analyzed by flow cytometry. Graphs show the number of $7AAD^-$ Lin^- $CD45^+$ cells (n=2, with 5-12 technical replicates). Lymphoid conditions; IL-7 and Flt3l (1 ng/ml each). Myeloid conditions; GM-CSF, M-CSF, IL-6 (5 ng/ml each) and SCF (1 ng/ml). (D,E) One-way ANOVA: *, *P*-value <0.05

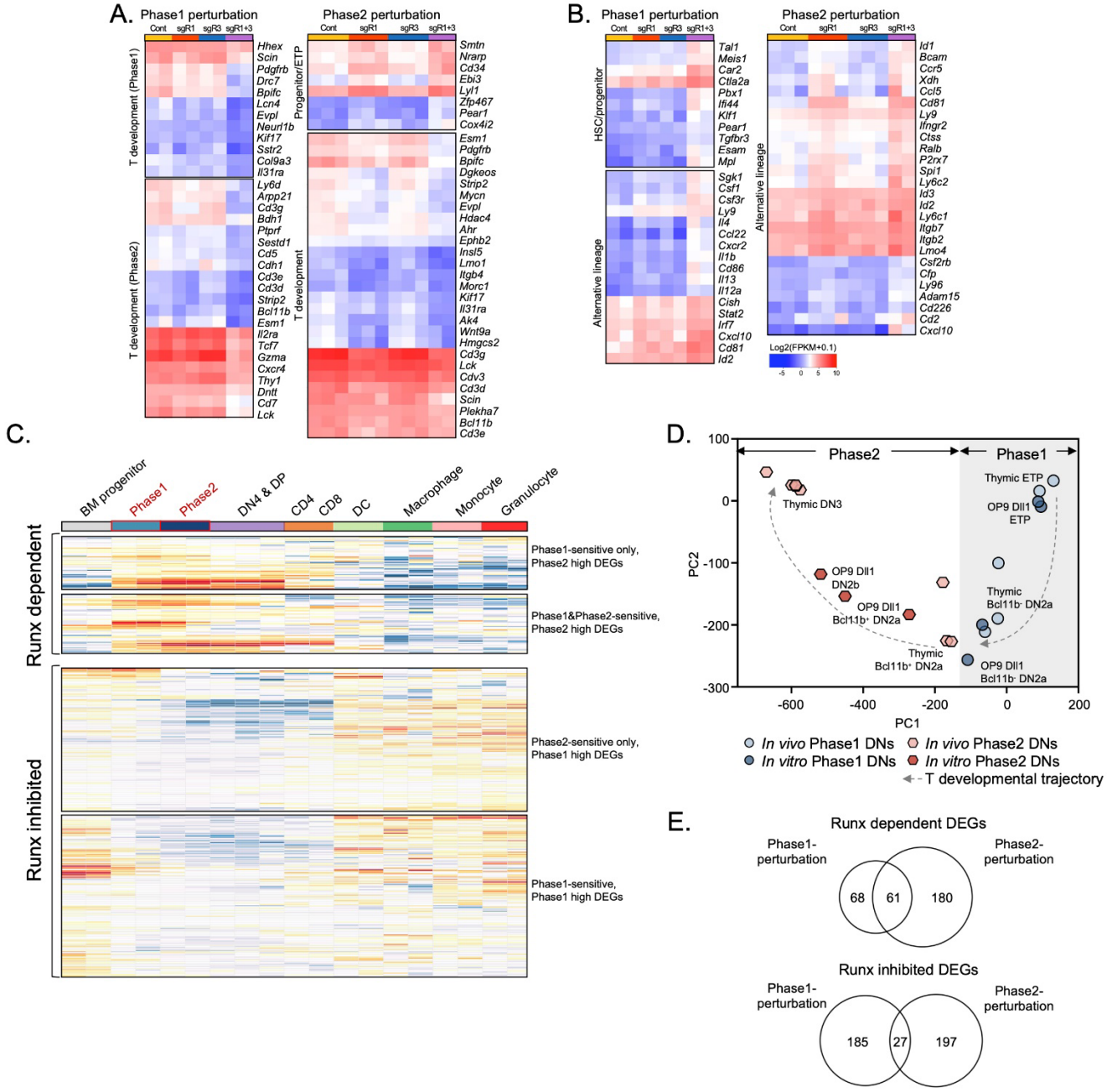

A.
Phase1 perturbation
Phase2 perturbation
Cont
sgR1
sgR3
sgR1+3
T development (Phase1)
T development (Phase2)
Progenitor/ETP
T development
Hhex
Scin
Pdgfrb
Drc7
Bpifc
Lcn4
Evpl
Neurl1b
Kif17
Sstr2
Col9a3
Il31ra
Ly6d
Arpp21
Cd3g
Bdh1
Ptprf
Sestd1
Cd5
Cdh1
Cd3e
Cd3d
Strip2
Bcl11b
Esm1
Il2ra
Tcf7
Gzma
Cxcr4
Thy1
Dntt
Cd7
Lck
Smtn
Nrarp
Cd34
Ebi3
Lyl1
Zfp467
Pear1
Cox4i2
Esm1
Pdgfrb
Bpifc
Dgkeos
Strip2
Mycn
Evpl
Hdac4
Ahr
Ephb2
Insl5
Lmo1
Itgb4
Morc1
Kif17
Il31ra
Ak4
Wnt9a
Hmgcs2
Cd3g
Lck
Cdv3
Cd3d
Scin
Plekha7
Bcl11b
Cd3e
B.
Phase1 perturbation
Phase2 perturbation
HSC/progenitor
Alternative lineage
Tal1
Meis1
Car2
Ctla2a
Pbx1
Ifi44
Klf1
Pear1
Tgfbr3
Esam
Mpl
Sgk1
Csf1
Csf3r
Ly9
Il4
Ccl22
Cxcr2
Il1b
Cd86
Il13
Il12a
Cish
Stat2
Irf7
Cxcl10
Cd81
Id2
Id1
Bcam
Ccr5
Xdh
Ccl5
Cd81
Ly9
Ifngr2
Ctss
Ralb
P2rx7
Spi1
Ly6c2
Id3
Id2
Ly6c1
Itgb7
Itgb2
Lmo4
Csf2rb
Cfp
Ly96
Adam15
Cd226
Cd2
Cxcl10
Log2(FPKM+0.1)
-5
0
5
10
C.
BM progenitor
Phase1
Phase2
DN4 & DP
CD4
CD8
DC
Macrophage
Monocyte
Granulocyte
Runx dependent
Runx inhibited
Phase1-sensitive only, Phase2 high DEGs
Phase1&Phase2-sensitive, Phase2 high DEGs
Phase2-sensitive only, Phase1 high DEGs
Phase1-sensitive, Phase1 high DEGs
D.
Phase2
Phase1
Thymic ETP
Thymic DN3
OP9 Dll1 ETP
OP9 Dll1 DN2b
OP9 Dll1 Bcl11b+ DN2a
Thymic Bcl11b- DN2a
Thymic Bcl11b+ DN2a
OP9 Dll1 Bcl11b- DN2a
PC1
PC2
100
0
-100
-200
-300
-600
-400
-200
0
200
In vivo Phase1 DNs
In vivo Phase2 DNs
In vitro Phase1 DNs
In vitro Phase2 DNs
T developmental trajectory
E.
Runx dependent DEGs
Phase1-perturbation
Phase2-perturbation
68
61
180
Runx inhibited DEGs
Phase1-perturbation
Phase2-perturbation
185
27
197

FIG. S4

Fig. S4. Runx1 and Runx3 support T-cell developmental pathways.

(A,B) Heatmaps show expression levels $\log_2$(FPKM+0.1) of the genes associated with (A) T development, or (B) alternative hematopoietic lineage and progenitor cells. (C) Heatmaps illustrate natural levels of expression patterns of Runx inhibited and Runx dependent genes in indicated immune cell populations using the ImmGen ULI RNA-seq dataset (www.immgen.org)(21). (D) Comparison of key transcriptome features of fresh *in vivo*-derived control thymocyte populations vs. *in vitro-derived* control populations from this study. Supervised PCA plot of data from RNA-seq is shown. Fixed PC loadings were determined based on single cell RNA-seq data (12) as described in Methods. (E) Area proportional Venn diagrams display the numbers of Runx dependent- and Runx inhibited-DEGs responding to Runx perturbation in Phase1 and Phase2.

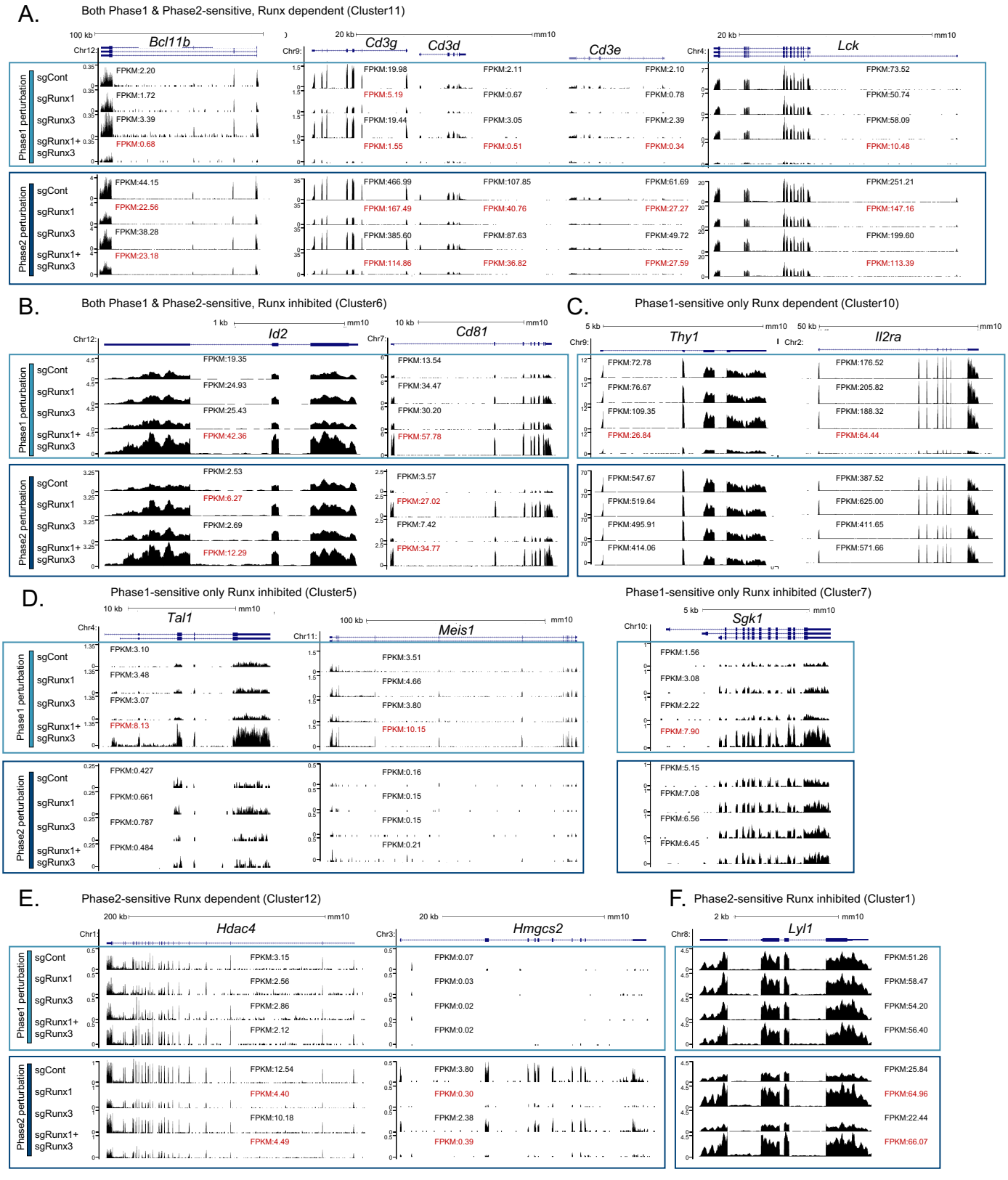
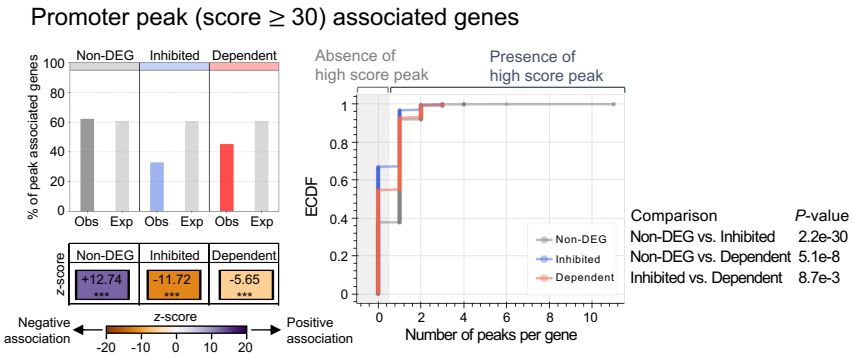

A.
Both Phase1 & Phase2-sensitive, Runx dependent (Cluster11)
Bcl11b
Cd3g
Cd3d
Cd3e
Lck
Phase1 perturbation
Phase2 perturbation
sgCont
sgRunx1
sgRunx3
sgRunx1+ sgRunx3
B.
Both Phase1 & Phase2-sensitive, Runx inhibited (Cluster6)
Id2
Cd81
C.
Phase1-sensitive only Runx dependent (Cluster10)
Thy1
Il2ra
D.
Phase1-sensitive only Runx inhibited (Cluster5)
Tal1
Meis1
Phase1-sensitive only Runx inhibited (Cluster7)
Sgk1
E.
Phase2-sensitive Runx dependent (Cluster12)
Hdac4
Hmgcs2
F.
Phase2-sensitive Runx inhibited (Cluster1)
Lyl1

Figure S5.

Fig. S5. Runx1 and Runx3 regulate target genes in a stage-sensitive manner.

Representative UCSC genome browser tracks for DEGs that are (A) both Phase1-sensitive and Phase2-sensitive Runx dependent, (B) both Phase1-sensitive and Phase2-sensitive Runx inhibited, (C) Phase1-sensitive only Runx dependent, (D) Phase1-sensitive only Runx inhibited, (E) Phase2-sensitive only Runx dependent, and (F) Phase2-sensitive only Runx inhibited are shown with FPKM values. Note that all tracks have same y axis values within the track group for a given phase, but some genes required a change of scale between their Phase1 and Phase2 track group.



Comparison	*P*-value
Non-DEG vs. Inhibited	2.2e-30
Non-DEG vs. Dependent	5.1e-8
Inhibited vs. Dependent	8.7e-3

Figure S6.

Fig. S6. Runx factor binding at promoter regions does not correlate with functional responsiveness of Runx target genes.

(A) The percentages of the genes in the indicated categories that are associated with Runx binding peaks are shown.Bar graphs compare observed values (Obs) with expected values (Exp) if binding among genomically expressed loci were random. Note that for both inhibited and dependent DEGs, there is actually less Runx binding at promoter sites than expected randomly or seen among non-DEGs. The color-map (below bar graph) represents the Z-scores (standardized residuals) across categories. (B) Cumulative frequency of the number of high score promoter peaks found per transcript in indicated group is shown. Despite measurable *P*-values, no subset of functional Runx targets is detected to show preferential binding.

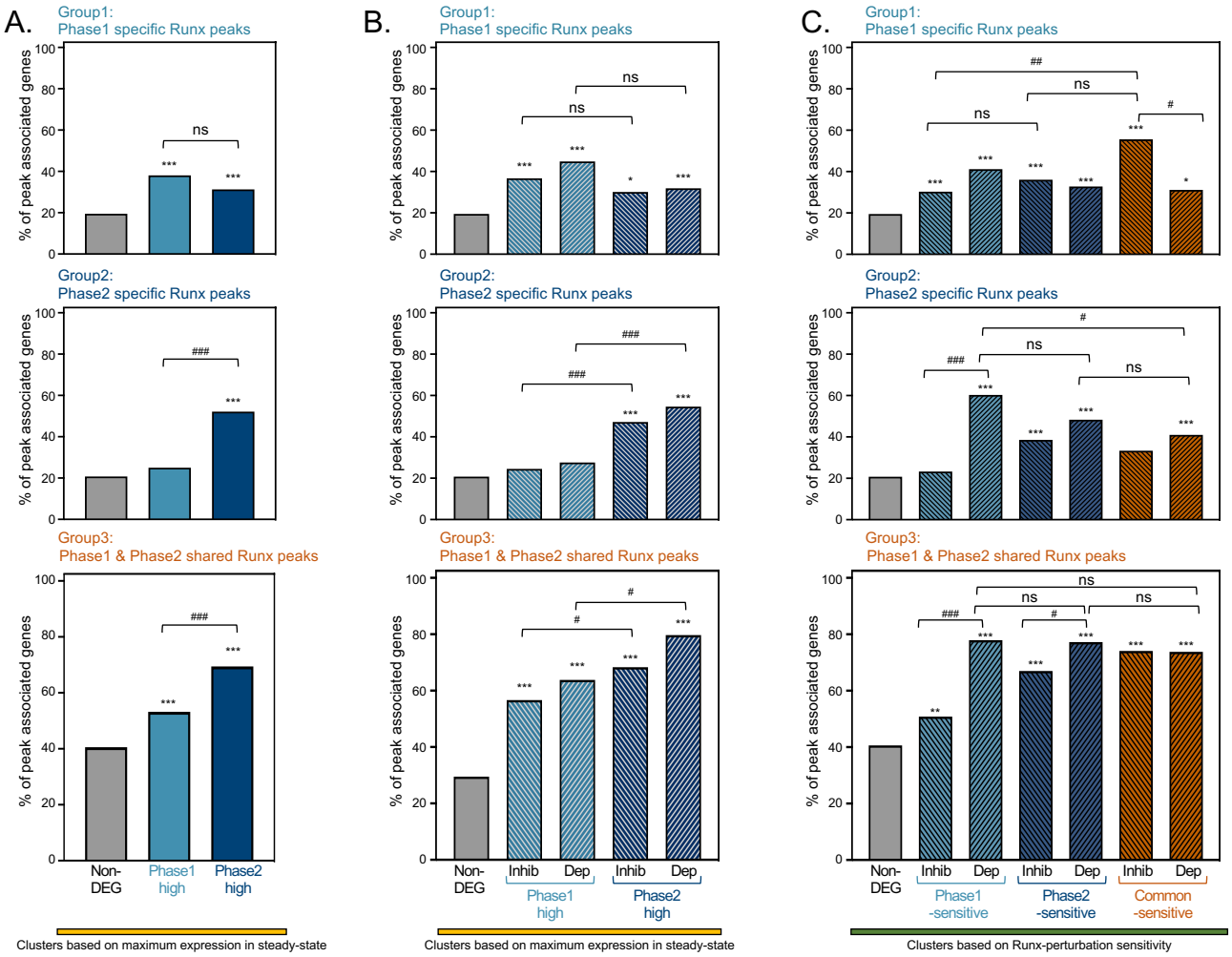
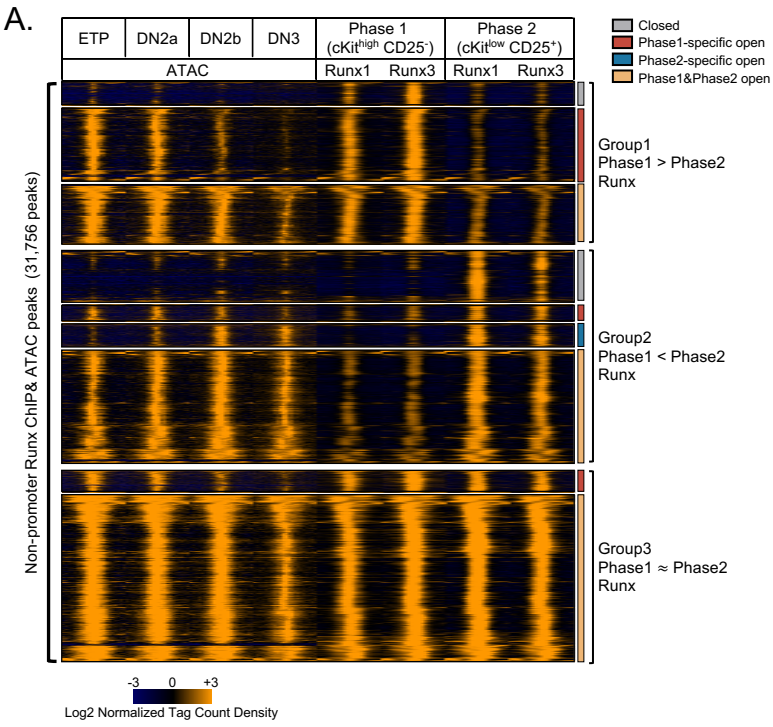

A.
Group1:
Phase1 specific Runx peaks
Group2:
Phase2 specific Runx peaks
Group3:
Phase1 & Phase2 shared Runx peaks
% of peak associated genes
Non-DEG
Phase1 high
Phase2 high
Clusters based on maximum expression in steady-state
B.
Group1:
Phase1 specific Runx peaks
Group2:
Phase2 specific Runx peaks
Group3:
Phase1 & Phase2 shared Runx peaks
% of peak associated genes
Non-DEG
Inhib
Dep
Inhib
Dep
Phase1 high
Phase2 high
Clusters based on maximum expression in steady-state
C.
Group1:
Phase1 specific Runx peaks
Group2:
Phase2 specific Runx peaks
Group3:
Phase1 & Phase2 shared Runx peaks
% of peak associated genes
Non-DEG
Inhib
Dep
Inhib
Dep
Inhib
Dep
Phase1 -sensitive
Phase2 -sensitive
Common -sensitive
Clusters based on Runx-perturbation sensitivity

Figure S7.

Fig. S7. Genomic binding of Runx factors is highly enriched at Runx functional target genes.

(A-C) Frequencies of Runx binding peaks of each occupancy group (Group 1, Group 2, Group 3) among DEGs with different expression characteristics. Category axes show sets of genes based on their responses to Runx perturbations and natural developmental patterns of expression. Y axes show the percentages of genes in those categories that are associated with at least one Runx-binding peak of the indicated type. Group1 peaks: Phase-1 specific occupancy. Group2 peaks: Phase2-specific occupancy. Group3 peaks: sustained occupancy across Phase1 and Phase2. (A) Comparison of peak type representation among clusters of genes divided simply between non-DEGs, DEGs with highest natural expression in Phase1, and DEGs with highest natural expression in Phase 2. (B) Comparison of peak type representation among clusters of genes divided as in A, but within each group further separating Runx-inhibited (inhib) from Runx-dependent (Dep) DEGs. (C) Comparison of peak type representations among clusters of genes separated according to the developmental stages when they are sensitive to Runx perturbation (regardless of their times of highest expression), as well as separating Runx-inhibited from Runx-dependent DEGs. *Z*-test. *, *P*-value <0.05, **, *P*-value<0.01, ***, *P*-value<0.001(comparison with non-DEG). Also, within indicated pairwise comparisons, $^{\#}$, *P*-value <0.05, $^{\#\#}$, *P*-value<0.01, $^{\#\#\#}$, *P*-value<0.001. For more finely grained distinctions (Fig. 6E), see Figs. S9 and S10 in this SI Appendix. Note two caveats in these analyses. Individual target genes were often linked to multiple sites of different types. Also, because ChIP-seq was carried out on pure Phase1 or Phase2 cells, whereas DEGs were scored three days after perturbation in Phase1 or Phase2, cells transitioning after perturbation contributed to the RNA data more than to the ChIP-seq data.


A.
ETP
DN2a
DN2b
DN3
Phase 1 (cKit^high CD25^-)
Phase 2 (cKit^low CD25^+)
ATAC
Runx1
Runx3
Runx1
Runx3
Closed
Phase1-specific open
Phase2-specific open
Phase1&Phase2 open
Non-promoter Runx ChIP& ATAC peaks (31,756 peaks)
Group1
Phase1 > Phase2
Runx
Group2
Phase1 < Phase2
Runx
Group3
Phase1 ≈ Phase2
Runx
-3
0
+3
Log2 Normalized Tag Count Density

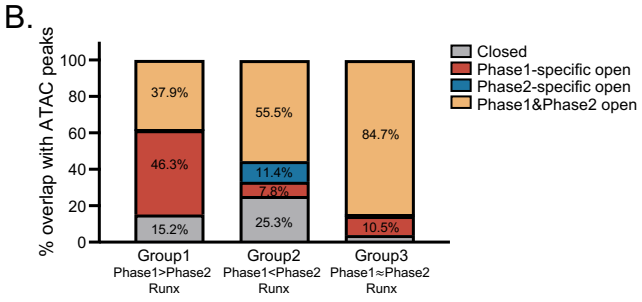
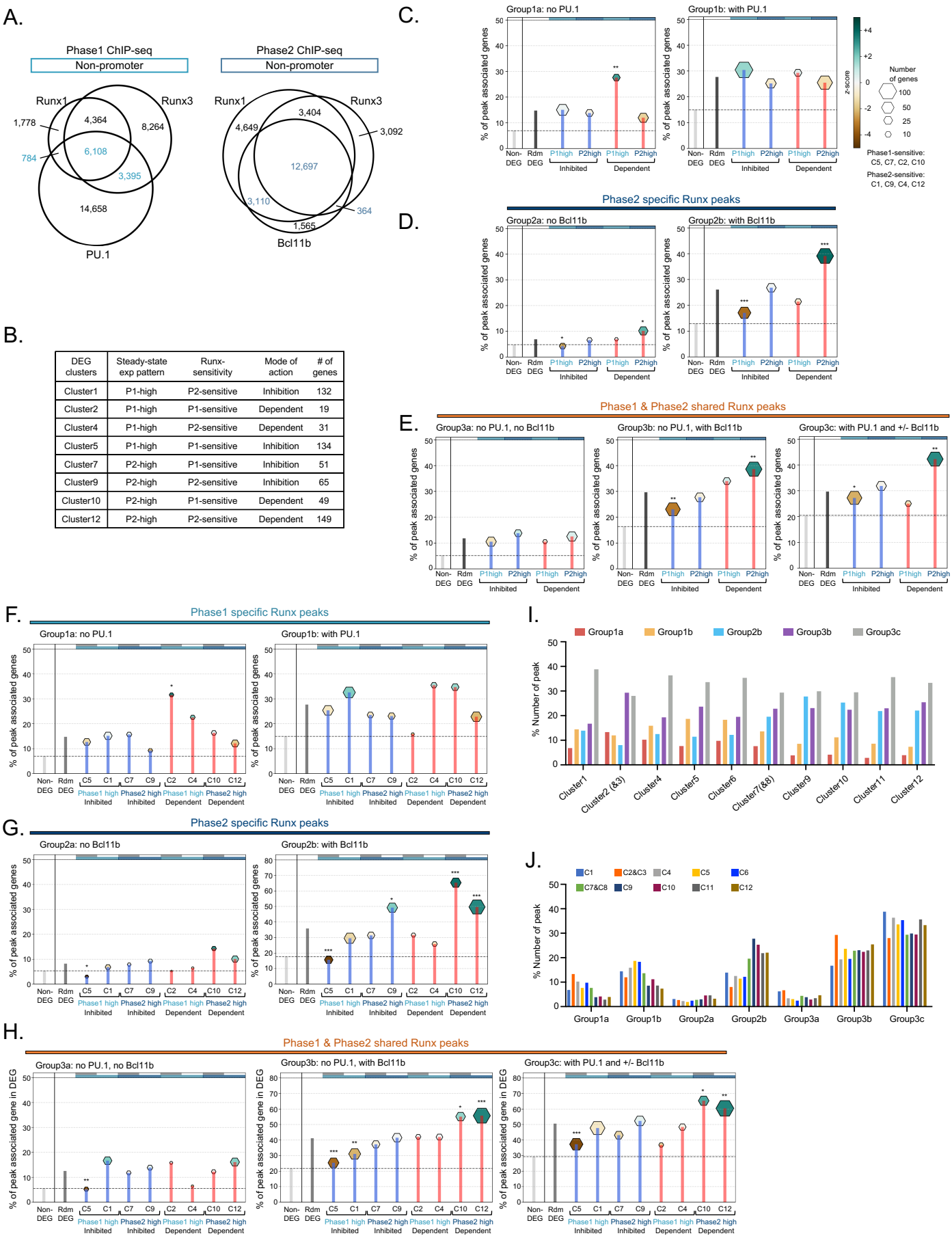

B.
% overlap with ATAC peaks
100
80
60
40
20
0
37.9%
46.3%
15.2%
55.5%
11.4%
7.8%
25.3%
84.7%
10.5%
Group1
Phase1>Phase2
Runx
Group2
Phase1<Phase2
Runx
Group3
Phase1≈Phase2
Runx
Closed
Phase1-specific open
Phase2-specific open
Phase1&Phase2 open

Figure S8.

Fig. S8. Chromatin accessibility change between Phase1 and Phase2 is distinct between different groups of Runx binding sites.

(A) ATAC sites and Runx1 and Runx3 bindings in pro-T cells are displayed by peak-centered heatmap. Data for ATAC-seq come from ImmGen (21). Tag count distributions are log_2 normalized. (B) Stacked column graph shows the percent of each class of ATAC sites within Group1, Group2, and Group3 Runx binding sites.



DEG clusters	Steady-state exp pattern	Runx-sensitivity	Mode of action	# of genes
Cluster1	P1-high	P2-sensitive	Inhibition	132
Cluster2	P1-high	P1-sensitive	Dependent	19
Cluster4	P1-high	P2-sensitive	Dependent	31
Cluster5	P1-high	P1-sensitive	Inhibition	134
Cluster7	P2-high	P1-sensitive	Inhibition	51
Cluster9	P2-high	P2-sensitive	Inhibition	65
Cluster10	P2-high	P1-sensitive	Dependent	49
Cluster12	P2-high	P2-sensitive	Dependent	149

Figure S9.

Fig. S9. Characterization of Runx binding site "Group" associations with DEG response type "Cluster" in Phase1 and Phase2 gene regulation.

(A) Area-proportional Venn diagrams display the peak overlap of Runx factors with PU.1 (Phase1) and Bcl11b (Phase2). The number of peaks is reported. (B) RNA-seq DEG clusters of interest are listed with descriptions and the number of genes.

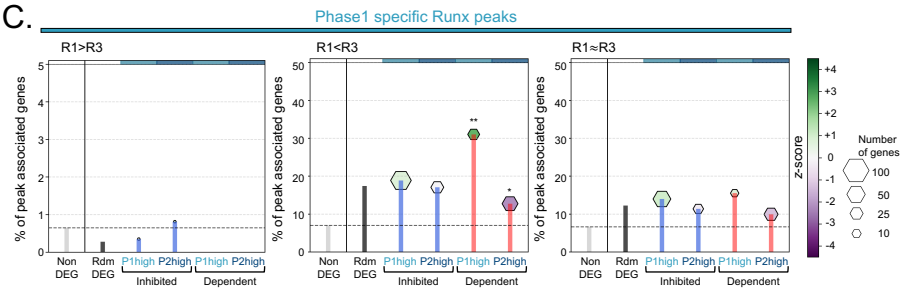
(C-H) Enrichment of specific site types is graphed as found in linkage to DEGs with particular response types. Panels C-E: DEG types subdivided broadly, as shown in Fig. S7B in this SI Appendix. Panels F-J: DEG types subdivided finely, as shown in panel B (& Fig. 6E). Percentages of genes in each Runx response cluster with linkage to Runx binding peaks of each Group are shown (y axes), compared with the percentage of all expressed non-DEGs (Non-DEG, adj *P*-value ≥ 0.05, average FPKM ≥ 1; dashed horizontal line) linked to that Group of binding peak. Specific associations are also tested for enrichment among DEGs by comparison with the percentage expected for that site Group if uniformly distributed among all DEGs (Rdm DEG). P1: Phase1; P2: Phase2. The percentages of the genes associated with such peaks (height of the spike) and the number of genes in DEG groups (size of hexagon) are plotted. Color map depicts *z*-scores (standardized residuals), calculated for relative enrichment of a given association among the DEG groups, i.e. relative to "Rdm DEG" (expected value of each DEG cluster), following Fisher's exact test. Dark teal, percentage in category strongly enriched over random. Dark brown, percentage in category strongly depleted below random. *, *P*-value <0.05, **, *P*-value<0.01, ***, *P*-value<0.001. (C-E) Runx peaks for broadly classified DEGs. (C) Phase1 specific peaks, (D) Phase2 specific peaks, and (E) Phase1 and Phase2 shared peaks. (F-H) Runx peaks linked to sub-classified DEGs, based on Clusters shown in Fig. 6E and Fig. S7B in this SI Appendix. (F) Phase1 specific peaks; (G) Phase2 specific peaks. (H) Phase1 and Phase2 shared peaks.

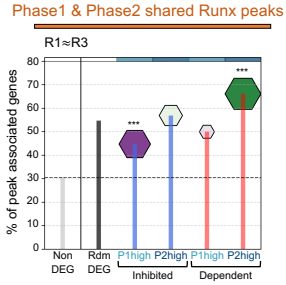
(I, J) Overall distribution of Runx peaks of each group among genes in each DEG cluster. Peaks of each Group were summed across all the genes of each DEG Cluster, and then plotted as the percentage they represent of total Runx peaks associated with the genes in that Cluster. Y axes in

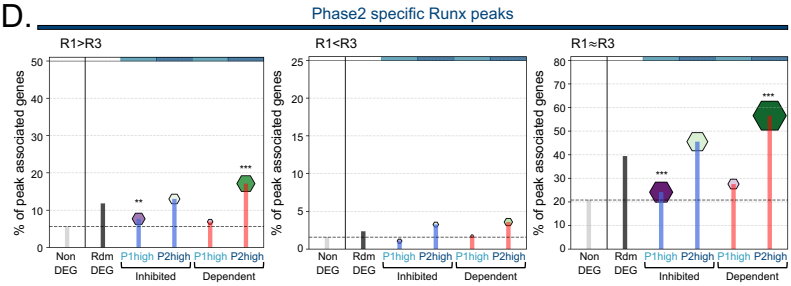
both graphs: number of peaks of the indicated Group in the indicated DEG Cluster as a percentage of all Runx peaks associated with genes in that DEG Cluster. (I) comparisons of peak Group representations among different DEG Clusters (sum for each Cluster ≈100%, but Groups 2a and 3a are not shown because of their very low values). (J) comparison of fractional representations within DEG Clusters among different peak Groups. Note shift from Phase1 expression-biased Clusters in Group 1a to Phase2 expression-biased clusters in Group 2b.

A. High score non-promoter peak groups

Groups		# of Peaks	Phase1		Phase2	
			Runx1	Runx3	Runx1	Runx3
Phase1 specific	R1>R3	64	+++	+	–	–
	R1<R3	943	+	+++	–	–
	R1≈R3	772	+++	+++	–	–
Phase2 specific	R1>R3	667	–	–	+++	+
	R1<R3	167	–	–	+	+++
	R1≈ R3	3,333	–	–	+++	+++
Phase1&2 shared	R1≈ R3	5,314	+++		+++	







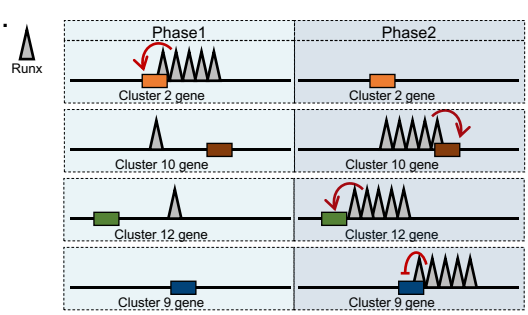


Figure S10.

Fig. S10. Differential functional impacts of Runx1 and Runx3 in target gene regulation in Phase1 and Phase2.

(A-D) Genomic interaction sites with different preferences for Runx1 vs. Runx3 binding and their relationships with Runx-regulated DEGs are analyzed. ChIP-seq data was categorized by strength of Runx1 and/or Runx3 genomic interactions in Phase1 and/or Phase2. Classification strategy and the number of peaks are shown (A). The percentage of the peak associated genes in broadly classified DEGs are displayed. Phase1 and Phase2 shared peaks (B), Phase1 specific peaks (C), and Phase2 specific peaks (D). (E) Diagram depicts features of Runx binding patterns in Phase1 and Phase2 around genes in indicated Clusters.

Figure S11.

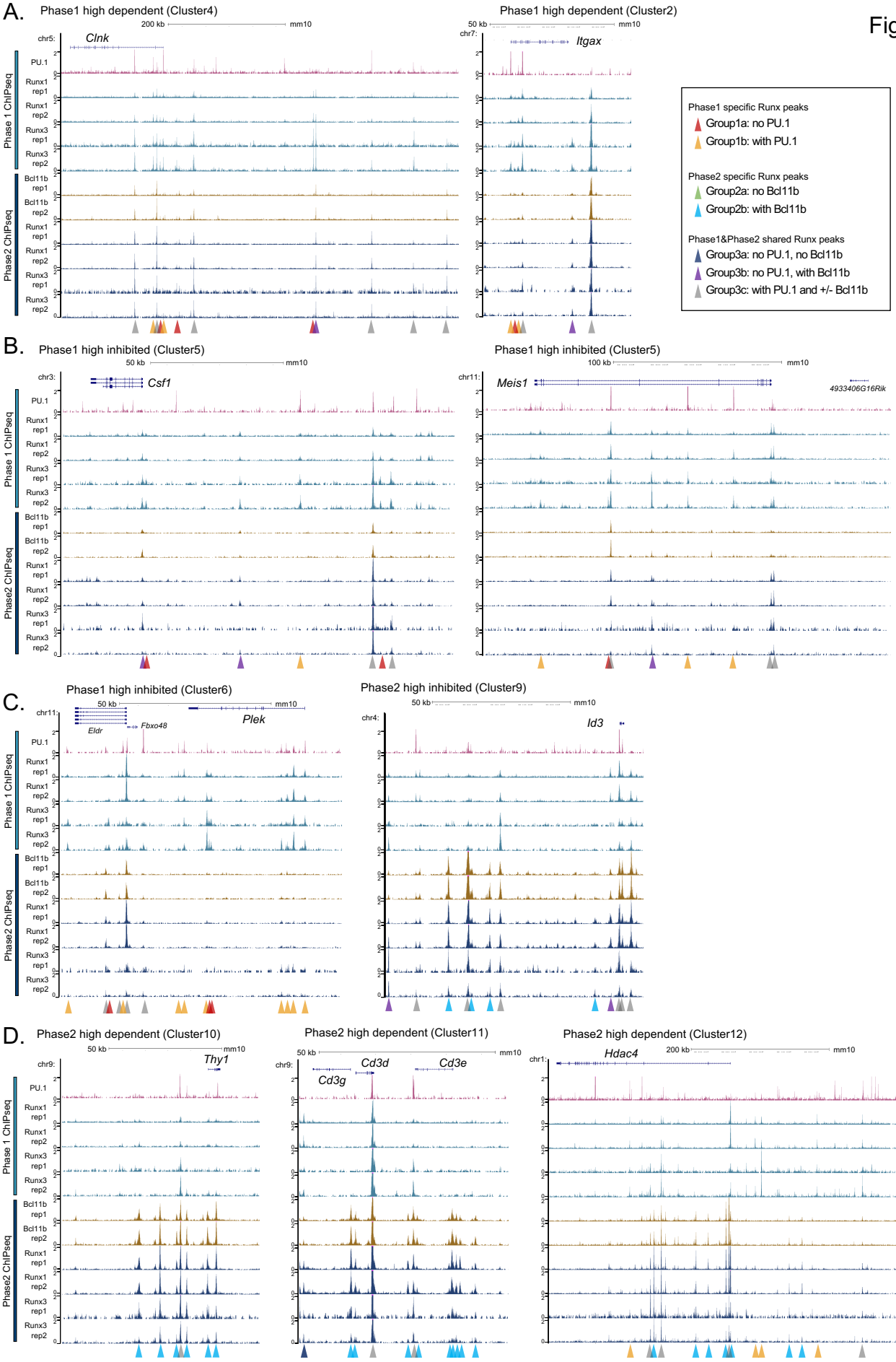

A.
Phase1 high dependent (Cluster4)
Clnk
Phase1 high dependent (Cluster2)
Itgax
Phase1 specific Runx peaks
Group1a: no PU.1
Group1b: with PU.1
Phase2 specific Runx peaks
Group2a: no Bcl11b
Group2b: with Bcl11b
Phase1&Phase2 shared Runx peaks
Group3a: no PU.1, no Bcl11b
Group3b: no PU.1, with Bcl11b
Group3c: with PU.1 and +/- Bcl11b
B.
Phase1 high inhibited (Cluster5)
Csf1
Phase1 high inhibited (Cluster5)
Meis1
4933406G16Rik
C.
Phase1 high inhibited (Cluster6)
Eldr
Fbxo48
Plek
Phase2 high inhibited (Cluster9)
Id3
D.
Phase2 high dependent (Cluster10)
Thy1
Phase2 high dependent (Cluster11)
Cd3g
Cd3d
Cd3e
Phase2 high dependent (Cluster12)
Hdac4
Phase 1 ChIPseq
Phase2 ChIPseq
PU.1
Runx1 rep1
Runx1 rep2
Runx3 rep1
Runx3 rep2
Bcl11b rep1
Bcl11b rep2

Fig. S11. Stage-sensitive Runx target genes possess phase-specific Runx binding sites.

Representative UCSC genome browser tracks surrounding indicated Runx DEG loci are displayed. Peaks with different stage-specificities and different associations with PU.1 and/or Bcl11b are highlighted with colored arrows according to the key. Representative tracks for genomic regions are shown near (A) Phase1 high Runx dependent genes (*Clnk*, *Itgax*); (B) Phase1 high Runx inhibited genes (*Csf1* and *Meis1*); (C) Phase1 high Runx inhibited genes (*Plek*) and Phase2 high Runx inhibited genes (*Id3*); and (D) Phase2 high Runx dependent genes (*Thy1*, *Cd3* clusters, and *Hdac4*).

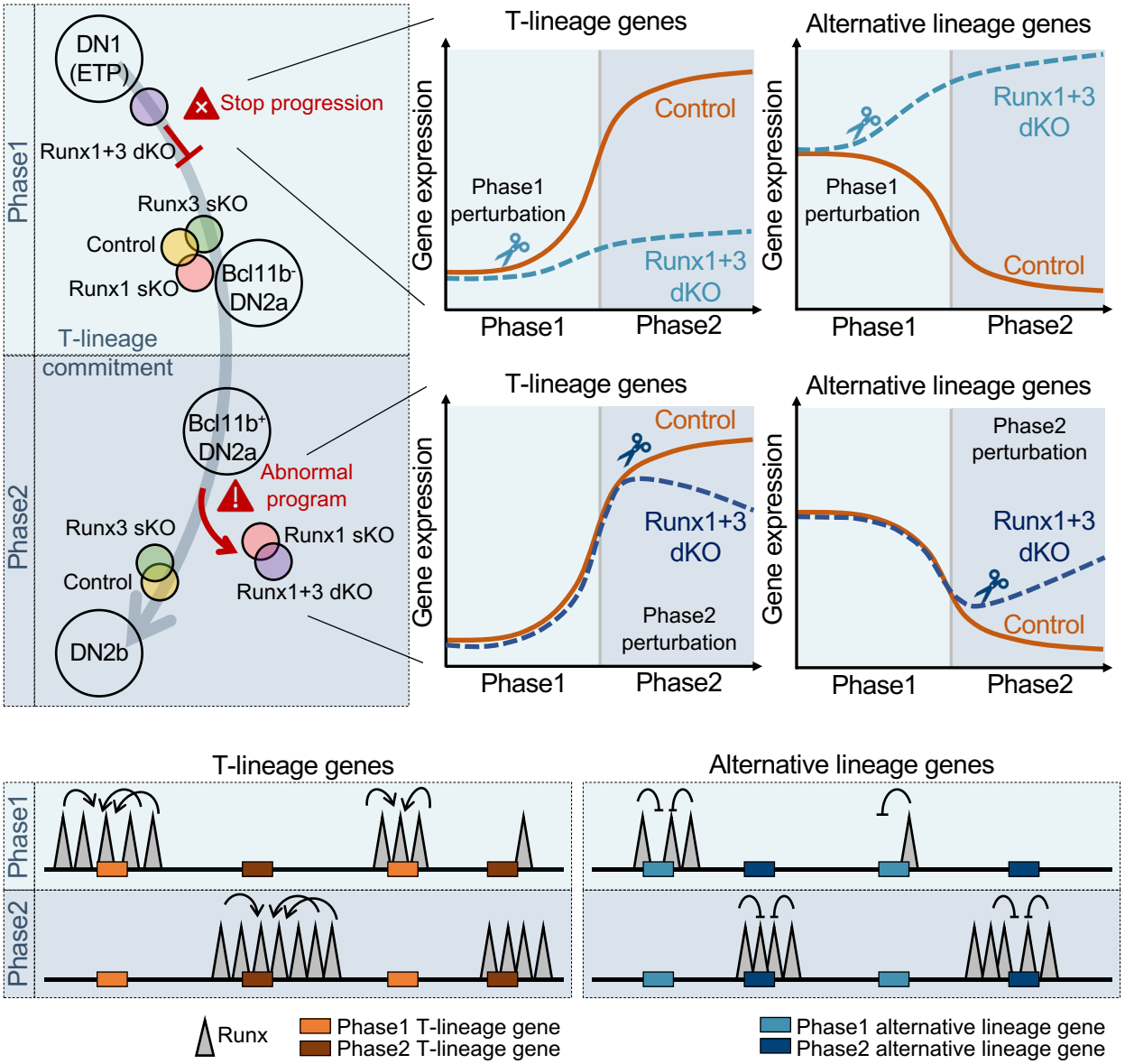

Phase1
DN1
(ETP)
Stop progression
Runx1+3 dKO
Runx3 sKO
Control
Runx1 sKO
Bcl11b-
DN2a
T-lineage
commitment
Phase2
Bcl11b+
DN2a
Abnormal
program
Runx3 sKO
Runx1 sKO
Control
Runx1+3 dKO
DN2b
T-lineage genes
Gene expression
Control
Phase1
perturbation
Runx1+3
dKO
Phase1
Phase2
Alternative lineage genes
Gene expression
Runx1+3
dKO
Phase1
perturbation
Control
Phase1
Phase2
T-lineage genes
Gene expression
Control
Runx1+3
dKO
Phase2
perturbation
Phase1
Phase2
Alternative lineage genes
Gene expression
Phase2
perturbation
Runx1+3
dKO
Control
Phase1
Phase2
T-lineage genes
Alternative lineage genes
Phase1
Phase2
Runx
Phase1 T-lineage gene
Phase2 T-lineage gene
Phase1 alternative lineage gene
Phase2 alternative lineage gene

Figure S12.

Fig. S12. Runx1 and Runx3 drive progenitor to T-lineage transcriptome conversion in mouse T-cell commitment via dynamic genomic site switching.

Graphical summary of this study.